

Assessing Public Awareness of Social Justice Documentary Films based on News Coverage versus Social Media

Jana Diesner¹, Rezvaneh Rezapour¹, Ming Jiang¹

¹University of Illinois at Urbana Champaign

Abstract

The comprehensive measurement of the impact that information products have on individuals, groups and society is of practical relevance to many actors, including philanthropic funding organizations. In this paper we focus on assessing one dimension of impact, namely public awareness, which we conceptualize as the amount and substance of attention that information products gain from the press and social media. We are looking at a type of products that philanthropic organizations fund, namely social justice documentaries. Using topic modeling as a text summarization technique, we find that films from certain domains, such as “Politics and Government” and “Environment and Nature,” attract more attention than productions on others, such as “Gender and Ethnicity”. We also observe that film-related public discourse on social media (Facebook and non-expert reviews) has a higher overlap with the content of a film than press coverage of films does. This is partially due to the fact that social media users focus more on the topics of a production whereas the press pays strong attention to cinematographic and related features.

Keywords: Impact assessment; natural language processing; topic modeling

doi: 10.9776/16305

Copyright: Copyright is held by the authors.

Acknowledgements: This work was supported by the FORD Foundation, grants 0125-6162, 0145-0558 and 0155- 0370, and by a faculty fellowship from the National Center of Supercomputing Applications (NCSA) at UIUC. We also thank Dr. Susie Pak from St. John’s University for her feedback and assistance.

Contact: (jdiesner, rezapou2, mjiang17)@illinois.edu

1 INTRODUCTION

As information science scholars, we often work on problems of societal relevance, such as the availability, governance and understanding of information. As part of this process, we ask what type and magnitude of impact certain types and pieces of information have on individuals, communities and society. Driving this point home, we can start by trying to understand the impact of our own research. With cross-fertilization from bibliometrics, straightforward ways for measuring the dissemination of scholarly publications have become accepted standards and common practice: we count the number of times that a paper gets cited, and compute metrics such as the h-index over these counts (Bornmann & Daniel, 2005; Hirsch, 2005). Over the last few years, various initiatives have emerged that aim to broaden this conceptualization and to consider the impact of scholarly work on society beyond academic boundaries. For example, altmetrics takes a wider range of research outcomes and their diffusion into account, e.g. by considering the sharing of data and code, and the mentioning of scholarly work on social media (Piwowar, 2013; Priem, Taraborelli, Groth, & Neylon, 2010; Thelwall, Haustein, Larivière, & Sugimoto, 2013).

Analogous but unrelated to these efforts in academia, philanthropic foundations have also started to develop new ways for identifying the impact of the work they are sponsoring in order to better understand the social return on their investments and their compliance with their missions and mandates (details on that below) (Barrett & Leddy, 2008; John & James, 2011; Knight Foundation, 2011). Consequently, foundations increasingly require their grantees to provide impact assessment reports at the end of their funding periods – a trend that has not yet been picked up on by science funding bodies, though such a move is at least imaginable in the future. Some foundations have published guidelines for impact assessment (Barrett & Leddy, 2008; Clark & Abrash, 2011; Ford Foundation; Green & Patel, 2013), others have released sample case studies (Britdoc). In contrast to academia, foundations conceptualize impact as positive change, or as moving the needle on some social justice issue (Ford Foundation; Green & Patel, 2013). For example, the Gates Foundation aims to collaboratively “support the development of innovative solutions [...] that can trigger change on a broader scale” (Gates Foundation), and the Knight Foundation is driven by the belief “that democracy thrives when people and communities are informed and engaged” (Knight Foundation). These goals might resonate well with

information science scholars who strive to solve societal problems. However, for an individual or small team grantee supported by a philanthropic foundation, it might be a challenging to overwhelming task to evaluate the impact of their work in a reliable, systematic, comprehensive and cost-efficient fashion. To bridge this gap, foundations have started to partner with academia to develop practical impact assessment solutions. The work presented in this paper is part of a larger collaboration with the Ford Foundation's JustFilms division, where we focus on studying the impact of issue-focused information products, mainly social justice documentaries. For this purpose, we have previously developed, implemented, applied and evaluated a theoretically grounded set of impact indicators, and a related methodology and technology (ConText) (Diesner, 2014; Diesner, Kim, & Pak, 2014; Diesner & Rezapour, 2015), as explained in more detail in the background section.

For issue-focused documentaries, multiple sets of impact dimensions have been previously defined (for a review see Diesner & Rezapour, 2015). Many classification schemas contain (as subset of) the following possible outcomes: public awareness, (consumer) attitudes and behavior, corporate policy, political action, and engagement with a film's partners (this set of factors defined by Britdoc). In this paper, we focus on measuring the first of these dimensions, namely public awareness. More specifically, we address the following questions:

1. What picture of an information product would one gain based on news coverage versus social media data?
2. How much do these impressions overlap with the actual content of the information product?

In this paper, we discuss prior work on impact assessment of issue focused media, introduce our datasets and analysis methods, present our findings, and discuss limitation, implications and next steps.

2 BACKGROUND

In academia, impact assessment has been mainly advanced in the field of environmental studies (Becker, Harris, McLaughlin, & Nielsen, 2002; Burdge & Vanclay, 1996; Vanclay, 1999, 2006). There, researchers use a combination of technical and social sensor data, including surveys, to assess the status quo of an ecosystem and identify implications of changes. In the field of communication, scholars have studied the impact of media products on individual cognition and behavior; primarily through lab studies and surveys (Whiteman, 2004). Lastly, in the field of political science, analysts have been using surveys to assess the impact of various stimuli on changes in people's opinion and/ or behavior (Devlin-Foltz, Fagen, Reed, Medina, & Neiger, 2012). In short, conducting surveys is common in impact assessment studies across different fields. The problem is that surveys are limited in scalability, and they also confront challenges related to the reliability of self-reported data as well as privacy concerns (Bernard et al., 1990).

Practitioners working on identifying the impact of issue-focused media typically employ (a combination of) the following data collection and analysis techniques: First, they use easy-to-interpret quantitative metrics, e.g. how many people saw a film or visited a webpage. Second, they collect press coverage of films. Third, they conduct focus-group interviews with small sets of participants, where the participants fill out before/after exposure questionnaires (for example Wang, 2012). The latter two types of data, i.e. press articles and interviews, are usually analyzed in a fairly qualitative fashion, such as close reading techniques. However, they also lend themselves to text analysis techniques, which scale better to large amounts of text documents (Diesner & Rezapour, 2015). Pursuing text mining techniques for the purpose of impact assessment has been recommended, but practical application are lagging behind (Chattoo & Das, 2014; Napoli, 2014). Our work has been targeting this gap.

Recent joint efforts by philanthropic foundations and universities have pushed the envelope on impact assessment solutions, or at least improved plans for putting this goal into action (for a survey see Chattoo & Das, 2014). For example, the Gates Foundation - with additional support from the Knight Foundation and the Open Society Foundations - has started the Impact Project at the Norman Lear Center of USC's Annenberg School for Communication and Journalism (Clark & Aufderheide, 2009; Media Impact Project). This project has resulted in a survey of the status quo of analysis methods as well as a theoretical framework for what a comprehensive impact assessment system should entail (Napoli, 2014). However, innovation in this area is still heavily driven by tools and solutions from outside of academia, such as SPARKWISE,¹ the Participant Index,² and the Story Pilot.³ The work presented in this paper brings a solution from the information science field to the table.

¹ <http://sparkwi.se>

In a nutshell, we apply a combination of techniques from social network analysis and natural language processing to data collected from news and social media in order to measure the public awareness around and issue and an information product about an issue. The purpose with this step is to identify a) the perception of a theme before a film is released (baseline model), b) changes in the baseline from the release onwards, and c) the overlap of the film's content (ground truth) with the film's coverage in social and print media (for details see Diesner et al., 2014). In most cases, we have been able to work with the makers and/ or producers of a film to identify the issues on which their production is meant to have an impact. Moreover, some of them have generously shared with us the transcripts of their films or book. Previously, we have conducted impact assessment studies on over a dozen films and used feedback from our project partners and the media impact practitioner community to improve our methodology and technology. The continuous dialogue and interaction with these stakeholders helped us to gain invaluable subject matter knowledge and to learn from their feedback. Moreover, a series of meetings⁴ between the involved stakeholders has helped to collaboratively advance the status quo of impact assessment of media, develop best practices, and train end users in making computational solutions useful for their work.

3 DATA

We considered eleven films and one book, listed in Table 1, for this study. This selection was made based on our collaboration with Ford Foundation. The main theme of each film (listed in Table 1) was given to us by the authors or producers of the films, or other domain experts. We grouped the films by topical sub-categories of social justice issues (Table 1 also introduces abbreviations for film titles for further reference in this paper). Furthermore, Table 1 lists the main awards that each product won or was nominated for. We consider awards as a different or alternative metric for success that is based on different criteria than the ones we use, e.g. cinematographic features. We use this information to qualitatively compare our results to how well a film does in terms of awards. In the rest of this section, we first discuss the classification schema we use for sub-domains of social justice information products, and then the data we collected per category and film.

We reviewed three category schemas for social justice issues: one used by the Ford Foundation,⁵ one from POV,⁶ which is PBS's documentary branch, and one from the United Nation's Human Rights division⁷. We mixed and matched these schemas to create a schema that covers both the films we have been analyzing and those that we plan to work on in the future. Our resulting top level categories are: "Art and Culture", "Criminal Justice", "Environment & Nature", "Ethnicity & Gender", "Family & Society", "Health & Health Care", "Politics & Government", "Rights and Liberties", "War & Peace", "Youth", and "International and Immigration". These top level classes can be further divided into sub-categories. For example, the head category of "Ethnicity & Gender" can entail sub-categories such as "Race and Racism", "Women", "Ethnic Conflicts" etc.. While the top-level classes are meant to be exhaustive for our purposes (of course the world might see new types of issues emerge in the future), they are not exclusive as some films fit into multiple categories. For example, "Pray the Devil Back to Hell" addresses both "Ethnicity & Gender" and "War & Peace". To select a best fitting category for that film as shown in Table 1, we used our knowledge about the film's content: as the main theme of the film is the impact of war on women and the role of women in peacebuilding processes, we chose the "Ethnicity & Gender" category.

We collected text data from news coverage and social media platforms. For news articles, we used LexisNexis Academic. The data collection process entailed two steps: First, we generated two queries to search articles about 1) the theme of the information product before and after the release of the film or book, and 2) the information product itself. We identified suitable keywords for query 1) through discussions with the filmmaker, author or funder. For query 2) we used the title of the information product and the name of the director or author; the latter part for disambiguation purposes. We inspected the

2 <http://www.participantmedia.com/participant-index>

3 <https://storypilot.org> (available for beta testing from the Harmony Institute)

4 For example <http://mediaimpactfunders.org/media-impact-focus-assessing-the-impact-of-media/>, <https://beta.cironline.org/events/dissection-c-impact/>, <http://www.themediaconsortium.org/2014-annual-conference/program/>

5 <http://www.fordfoundation.org/issues>

6 <http://www.pbs.org/pov/discover/#.Veyl8xFVhBc>

7 <http://www.ohchr.org/EN/Issues/Pages/ListOfIssues.aspx>

retrieved documents for false positives and refined the queries accordingly through multiple iterations. Ultimately, we retrieved three sets of data per film from LexisNexis Academic: press on the theme before and after release, and press coverage of the film. In this paper, we focused on the analysis of press on film, but provide comparative frequency counts for all types of data (in Table 2).

	Name of product	Abbrev.	Main theme	Awards	Reference
Top level theme: Gender and Ethnicity					
1	Solar Mamas	SOMS	Women, Education in Developing Countries	<ul style="list-style-type: none"> •Won: Cinema Eye Honors Awards 2014, Sheffield Internat. Documentary Festival 2013 •Nominated: Munich International Documentary Festival 2013 	(Eldaief & Noujaim, 2012)
2	Peace Unveiled	PUV	Women, War, Peace Building, Women Rights, Afghanistan	<ul style="list-style-type: none"> •Won: Gracie Allen Awards 2012 (One subject in PDBTH (Leymah Gbowee) also won a Nobel Peace Prize after the release of the film in 2011) 	(Reticker, 2011b)
3	Pray The Devil Back to Hell	PDBTH	Women, War, Peace Building, Protest Against Charles Taylor, Liberia		(Reticker, 2008)
4	The War We Are Living	WWAL	Women, War, Peace Building, Gold Land, Colombia		(Reticker, 2011b)
5	I Came to Testify	ICTT	Women, War, Peace Building, Sexual Assault, Serbia		(Reticker, 2011a)
6	Through A Lens Darkly	TALD	African American, Photography, Racial Issues	<ul style="list-style-type: none"> •Won: African Movie Academy Awards 2014, Image Awards 2015, Los Angeles Pan African Film Festival 2014, Santa Barbara International Film Festival 2014 •Nominated: Black Reel Awards 2015 	(Harris, 2014)
Top level theme: Politics and Government					
7	The House I Live In	HILI	Mandatory Minimum Sentencing, Prison, Drugs	<ul style="list-style-type: none"> •Won: Golden Trailer Awards 2013, Silverdocs Documentary Festival 2012, Sundance Film Festival 2012 •Nominated: Houston Film Critics Society Awards '12, Detroit Film Critic Society '12 	(Jarecki, 2012)
8	Fed Up	FEDUP	Added Sugar and Sugar Tax	<ul style="list-style-type: none"> •Nominated: Sundance Film Festival 2014 	(Soechtig, 2014)
9	Pay 2 Play	P2P	Politics, USA, Election, Influence of Money	N/A	(Ennis, 2014)
Top level theme: Environment and Nature					
10	This Changes Everything	TCE	Climate Change, Economic Inequality, Capitalism	<ul style="list-style-type: none"> •Won: Hillary Weston Writers' Trust Prize for Nonfiction, New York Times non-fiction bestseller 	(Klein, 2014)
11	Pandora's Promise	PAPR	Nuclear Power and Sustainability	<ul style="list-style-type: none"> •Won: Sheffield International Documentary Festival 2013 	(Stone, 2013)
Top level theme: Family and Society					
12	One Mile Away	OMA	Gangs, Violence, Peace, England	<ul style="list-style-type: none"> •Won: Edinburgh International Film Festival 2012 	(Woolcock, 2012)

Table 1: Information products considered

For social media, we used the public Facebook fanpages of each production if available. Data collection was done by using NodeXL (Hansen, Shneiderman, & Smith, 2010). We considered posts and comments on the fanpages; excluding posts from the page administrator (for FedUp, we failed to collect Facebook data due to the huge number of data points that the importer struggled with). For non-expert

reviews, which are another type of social media data, we collected reviews from Amazon. Finally, our text data for this study also entailed the transcripts of the films.

The number of retrieved documents per film and source is shown in Table 2⁸. While the number of cases considered is small (12 information products), the number of press articles (N=19,675) and Facebook posts and comments (N=2,819) is sizable. Hence, with respect to social justice categories, our findings are based on a small sample size, which are therefore rather qualitative or indicative of trends, while the results with respect to text mining are based on larger samples, and therefore, they are more generalizable within the given domain and time frame.

Info Product	Press on film	Press on theme before	Press on theme after	Facebook Comments and Posts	Reviews
Gender and Ethnicity					
SOMS	72	1,418	1,317	103	6
PUV	4	450	1,069	47	0
PDBTH	85	493	605		46
WWAL	3	80	109		0
ICTT	22	54	66		0
TALD	24	1,517	1,570	140	9
Politics and Government					
HILI	110	135	309	832	211
FEDUP	41	1,668	865	0	1,217
P2P	1	206	256	206	16
Environment and Nature					
TCE	158	1,793	1,736	355	577
PAPR	57	1,005	754	860	41
Family and Society					
OMA	71	802	750	276	0

Table 2: Number of documents per information product and source

4 METHODS

In order to identify the gist of information from our text sets, we apply the same summarization technique to both types of data in order to identify a) the factual impression of an information product that one would gain based on news articles versus social media and b) how these impressions compare against each other and to the content of the film. More specifically, we used topic modeling; a text summarization technique that represents the gist of a body of information in terms of the distinct salient themes that are explicitly or implicitly contained in the data (Blei, 2012; Blei, Ng, & Jordan, 2003; Griffiths, Steyvers, & Tenenbaum, 2007). Each topic is represented as a vector of words, and the words are ordered by decreasing fit with the topic. Also, each topic has a fitness value that describes its strength of association with a given corpus. Topics are by default not labeled, though people occasionally use the first term per topic as a label. Topic modeling has become a highly popular technique in the computational social sciences and digital humanities because it is a fast, fully automated and non-deterministic technique. The main caveats with topic modeling are evaluation and interpretation of results, which is due to the unsupervised nature of this method (Chang, Boyd-Graber, Gerrish, Wang, & Blei, 2009).

We used the ConText tool for topic modeling, which leverages the Mallet for this routine (McCallum, 2002). Each model was generated multiple times and results were compared across runs to identify a somewhat robust result. In order to interpret and contrast the outcomes, two of the authors each individually and independently labeled each topic. Then, all three of us compared the results and resolved each disagreement or offset through discussion. The label set was also normalized through discussion for the sake of consistency. We find that this process requires a moderate to substantial amount of familiarity with both the content of an information product and the social justice topics considered.

⁸ The films “Peace Unveiled,” “Pray The Devil Back to Hell,” “The War We Are Living,” and “I Came to Testify” are part of a five-part series on Women, War and Peace (WWP), which share a Facebook fan page.

	Transcript	Press on film	Facebook	Reviews
SOMS	<ul style="list-style-type: none"> • 50% Family, Work, Women • 30% Training, Education • 17% Social Issues 	<ul style="list-style-type: none"> • 40% Social Issues, Women • 30% Film • 20% Film Festivals, Film Screening 	<ul style="list-style-type: none"> • 40% Film, Film Festivals, Social Issues • 40% Women, Producer, Documentary film • 20% Storytelling 	<ul style="list-style-type: none"> • 40% Solar Power • 30% Women • 25% Education, Film, Documentary
PUV	<ul style="list-style-type: none"> • 50% War, Women, Karzai, Afghanistan • 30% Clinton, Negotiate, Government • 20% Afghan, Peace, Work 	<ul style="list-style-type: none"> • 50% Women, War, Peace, Story • 20% Music, Brooklyn • 17% Media 	<ul style="list-style-type: none"> • 40% Nobel Prize, Violence • 30% Women, War, Peace, Producer • 20% L. Gbowee 	<ul style="list-style-type: none"> • N/A
PDBTH	<ul style="list-style-type: none"> • 35% Subject, Film, Peace, Liberia • 30% Campaign, Africa • 25% Women, War, Charles Taylor 	<ul style="list-style-type: none"> • 55% Women, War, Peace, Liberia, Producer • 30% Awards, Film Festivals • 20% Film, Screening 	<ul style="list-style-type: none"> • 40% Nobel Prize, Violence • 30% Women, War, Peace, Producer • 20% L. Gbowee 	<ul style="list-style-type: none"> • 60% Women, Film, Liberia • 35% War, Movie • 1% C. Taylor
WWAL	<ul style="list-style-type: none"> • 40% Colombia, Government • 20% War, Peace • 15% Gold, Women, Problem 	<ul style="list-style-type: none"> • 40% Women, War, Peace, Series • 40% Film Screening • 20% Film, Charles Taylor 	<ul style="list-style-type: none"> • 40% Nobel Prize, Violence • 30% Women, War, Peace, Producer • 20% L. Gbowee 	<ul style="list-style-type: none"> • N/A
ICTT	<ul style="list-style-type: none"> • 50% Witness, Rape, War, Women, Sexual Assault • 30% Crime, Yugoslavia, Slaughter, War Time • 15% Males 	<ul style="list-style-type: none"> • 30% Trial • 30% Film, Violence, Justice • 20% L. Gbowee • 20% Film Screening 	<ul style="list-style-type: none"> • 40% Nobel Prize, Violence • 30% Women, War, Peace, Producer • 20% L. Gbowee 	<ul style="list-style-type: none"> • N/A
TALD	<ul style="list-style-type: none"> • 50% African American, Photography • 50% Family 	<ul style="list-style-type: none"> • 60% Film, Documentary Film • 20% African American 	<ul style="list-style-type: none"> • 40% Film, Film Festivals, Documentary film • 30% Director • 30% Film Screening 	<ul style="list-style-type: none"> • 50% African American, Photography • 20% African American, People • 20% Film, Photo, Family
HILI	<ul style="list-style-type: none"> • 65% War on Drugs, Social Issues • 20% Prison, African American. • 15% Law, Public Debate 	<ul style="list-style-type: none"> • 60% Film, Documentary, Director • 10% Film Festivals • 25% Film Screening 	<ul style="list-style-type: none"> • 40% War On Drugs • 20% Drugs, Justice • 10% Prison 	<ul style="list-style-type: none"> • 50% War On Drugs, Prison • 40% Film, Documentary • 5% Director
FEDU _D	<ul style="list-style-type: none"> • 80% Junk Food • 10% Overweight People, Social Story 	<ul style="list-style-type: none"> • 65% Food Industry, Film, Documentary • 20% Film Making • 10% Film Screening 	N/A	<ul style="list-style-type: none"> • 40% Documentary, Film • 50% Junk Food • 10% Healthy Food
P2P	<ul style="list-style-type: none"> • 50% People, Money, Politics • 20% Election • 15% Personal Story on Topic 	<ul style="list-style-type: none"> • 60% Film, Documentary • 20% Influence of Film • 20% Movies, Politics 	<ul style="list-style-type: none"> • 30% Money, Politics • 25% Film Screening • 25% Film • 20% People 	<ul style="list-style-type: none"> • 50% Politics, Money, Film • 40% World, Money, Social Issues, Film
TCE	<ul style="list-style-type: none"> • 40% Energy Resources • 20% Companies • 30% Politics, Global Warming 	<ul style="list-style-type: none"> • 50% Book • 30% Capitalism, Climate Change • 15% Economic Growth 	<ul style="list-style-type: none"> • 35% Book, Author • 25% Climate Change • 20% Government • 20% Social Issues 	<ul style="list-style-type: none"> • 55% Book, Author • 25% Climate Change, Politics, Economy • 12% Global Warming
PAPR	<ul style="list-style-type: none"> • 60% Nuclear Power, Social Issues • 25% Safety • 15% China, Climate 	<ul style="list-style-type: none"> • 60% Nuclear Power, Film • 35% Documentary Film, Festival • 10% Screening 	<ul style="list-style-type: none"> • 45% Nuclear Power, Film • 20% Social Issues, Climate, Director • 20% Sustainable Energy 	<ul style="list-style-type: none"> • 65% Film, Documentary, Nuclear Power, Climate Change • 20% Energy, Growth • 15% Nuclear Power Fail
OMA	<ul style="list-style-type: none"> • 20% Men, Birmingham • 15% Killing • 15% School, Thieve 	<ul style="list-style-type: none"> • 60% Film, Documentary, Director • 10% Film Festivals • 25% Film Screening 	<ul style="list-style-type: none"> • 50% Film Screening • 30% Documentary, Work • 30% People, Change, Birmingham 	<ul style="list-style-type: none"> • N/A

Table 3: Summarization (topic modeling) and comparison of content of information product versus public discourse sources

5 RESULTS

The number of documents per source (Table 2) suggests that the amount of attention (as a proxy for social awareness) that a documentary attracts on media (press on film) versus social media does not necessarily correlate with the popularity of the topic of the film and vice versa (Table 2). For example, “*The House I Live in*”, a film about reforming mandatory minimum sentencing (MMS) for jail time for drug offenses, got more coverage from press on film, social media and reviews than regular media (press after film). This movie was the most impactful film when conceptualizing impact as frequent mentioning on a variety of channels. On the opposite end of the spectrum is “*Peace Unveiled*”, a film about the participation of women in peace negotiations with the Taliban in Afghanistan, where the topic itself got decent press coverage, but the film addressing the issue did not generate wide attention – at least on the channels we consider. We note that the comparison of the number of posts on Facebook to news articles and reviews is skewed in that Facebook comments are shorter and take less effort to be written.

The results in Table 2 also indicate that both types of social media data (reviews and Facebook comments), which are typically authored by laymen and hence provide an indicator of public attention from non-experts, are more numerous for films related to local issues that impact peoples’ daily lives and well-being (the categories of “Environment & Nature” and “Politics and Government”), and less copious for films on global issues that might seem more remote. For example, “*Fed Up*”, a film about health and politics, “*This Changes Everything*”, a book about environment and capitalism, and “*Pandora’s Promise*”, a film about nuclear power, have the largest number of documents from social media. One possible reason for this observation might be that people can relate more easily to themes with which they have first-hand experience. Press articles, which are typically written by professional journalists, are supposed to take a more neutral and comprehensive perspective, which could explain why we see a more balanced number of articles across sub-categories.

The topic modeling results (Table 3 and Table 4) show that press on film focuses more on addressing film screenings and portraying a documentary or book as an artistic or intellectual product than the transcript does. More specially, about half or more of the topics and key terms from press on film refer to artistic features, prize nominations and awards, directors and producers, and screening announcements. The remaining topics and terms typically relate to the actual issue raised in the film. Consequently, the overlap in content between the transcript and the press on film (Table 4) is low to medium; indicating a higher interest of the media in the art of film making than the addressed issues. In fact, the only film with a high overlap between transcript and press on film is “*Pray the Devil Back to Hell*”.

On the other hand, the results based on Facebook show a more balanced combination of issue-related information and artistic features, with the latter having a slightly lower presence than the former. This means that the user community discusses the issues at stake, and occasionally even introduces additional yet issue focused terms, like using “sustainable energy” instead of “nuclear power” (for the case of “*Pandora’s Promise*” (PAPR)). For Facebook, the overlap with the transcript’s content is higher than for press on the film.

Interestingly, the most influential films based on attention on Facebook (“*The House I Live In*,” “*Pandora’s Promise*”) as well as the book (“*This Changes Everything*”) also won or were nominated for the most prestigious awards, e.g. at Sundance (for a list of awards see Table 1). The press on film reflects this correlation less strongly.

Products	Overlap with Transcript		
	Press On Film	Facebook	Non-expert Reviews
SOMS	Low	Low	High
PUV	Medium	Low	N/A
PDBTH	High	Medium	High
WWAL	Low	Low	N/A
ICTT	Medium	Low	N/A
TALD	Low	Low	High
HILI	Low	High	Medium
FEDUP	Low	N/A	Medium
P2P	Low	Medium	Medium
TCE	Medium	High	Medium
PAPR	Medium	High	High
OMA	Low	Low	N/A

Table 4: Final overlap result of summarization

Finally, the result (Table 4) show that the non-expert reviews have higher overlap with the transcripts compared to the other two sources. This might be because some reviews start with a brief synopsis of the film. Overall, we find that social media (which includes Facebook comments and posts and reviews) have a higher overlap with the transcripts or content of a film than press coverage of a films does.

6 CONCLUSION, LIMITATIONS AND FUTURE WORK

Funders of issue-focused media have recently started to develop ways to assess the impact of information products in an empirical, systematic and comprehensive fashion. These efforts could matter to the academic community because science funders might at some point begin to require impact assessment plans as parts of grant proposals in the future, and academic hiring and promotion committees might start to take a data-driven look at a scholar's achievements beyond their citation counts. As information scientists, we can contribute substantive knowledge and skills for understanding and assessing the effectiveness and diffusion of information to these efforts.

One dimension of the wider societal impact of information products is the attention that these productions gain in the press and on social media. We herein focus on a single dimension of impact, i.e. the amount and substance of resonance that a film attracts on various public channels, namely traditional press and social media.

We observe that the domains of "Politics & Government" and "Environment & Nature" are associated with a) a higher volume of public attention and b) a stronger overlap in content between the transcript and both press on film and social media than the domains of "Ethnicity & Gender" and "Family & Society".

What picture of an information product would one gain based on news coverage versus social media data? Classic media devote a large amount of writing to topics related to filmmaking, while social media feature a more balanced attention to the substance of a film and its artistic features. How much do these impressions overlap with the actual content of the information product? Our findings suggest that press coverage of films has a lower intersection with the film's content than observed for social media data – which might be explained by the answer to the question above. This potentially missed opportunity can be addressed by informing outreach people on the film production site as well as journalists to tie the coverage of a film in the media to facets of the current public discourse on this issue. The summarization technique used herein can help practitioners with this step, even though topic modeling suffers from the limitations inherent to unsupervised techniques.

Finally, we observe an overlap between the productions that perform well based on our conceptualization of impact (i.e. public awareness for the case of this paper) and those that have been nominated for or have won major relevant awards. So far, this is a purely correlational effect. Both success metrics are based on vastly different features: awards consider films as holistic products that involve cinematography, sounds, other artistic elements, and its narrative and impact. We reduce films and books to the content of their transcripts - a crude step that filmmakers have previously criticized - and then observe the perception of a product on formal (press) and informal (social media) public discourse platforms. However, there might be underlying factors that drive impact into the same direction. We plan to scrutinize this alignment in more depth and utilize a larger sample for this purpose in our future research.

7 REFERENCES

- Barrett, D., & Leddy, S. (2008). *Assessing Creative Media's Social Impact*. Retrieved from
- Becker, D., R., Harris, C., C., McLaughlin, W., J., & Nielsen, E., A. (2002). A participatory approach to social impact assessment: the interactive community forum. *Environmental Impact Assessment Review*, 23(3), 367-382. doi:10.1016/S0195-9255(02)00098-7
- Bernard, H. R., Johnsen, E. C., Killworth, P. D., McCarty, C., Shelley, G. A., & Robinson, S. (1990). Comparing four different methods for measuring personal social networks. *Social Networks*, 12(3), 179-215.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993-1022.
- Bornmann, L., & Daniel, H.-D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3), 391-392.

- Britdoc. *Real films, real impact*. http://britdoc.org/real_good/evaluation
- Burdge, R., J., & Vanclay, F. (1996). Social impact assessment: a contribution to the state of the art series. *Impact Assessment*, 14(1), 59-86.
- Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009, December). *Reading tea leaves: how humans interpret topic models*. Paper presented at the Neural Information Processing Systems (NIPS), Vancouver, Canada.
- Chattoo, C. B., & Das, A. (2014). *Assessing the social impact of issues-focused documentaries: research methods & future considerations*. Retrieved from
- Clark, J., & Abrash, B. (2011). *Social justice documentary: designing for impact*. Retrieved from <http://www.centerforsocialmedia.org/designing-impact>
- Clark, J., & Aufderheide, P. (2009). *Public media 2.0: dynamic, engaged publics*. Retrieved from
- Devlin-Foltz, D., Fagen, M. C., Reed, E., Medina, R., & Neiger, B. L. (2012). Advocacy evaluation challenges and emerging trends. *Health promotion practice*, 13(5), 581-586.
- Diesner, J. (2014). *ConText: software for the integrated analysis of text data and network data*. Paper presented at the Conference of International Communication Association (ICA), Seattle, WA.
- Diesner, J., Kim, J., & Pak, S. (2014). Computational Impact Assessment of Social Justice Documentaries. *Journal of Electronic Publishing*, 17(3). doi:10.3998/3336451.0017.306
- Diesner, J., & Rezapour, R. (2015). *Social Computing for Impact Assessment of Social Change Projects*. Paper presented at the International Social Computing, Behavioral Modeling and Prediction Conference (SBP), Washington, DC.
- Eldaief, M., & Noujaim, J. (Writers). (2012). Solar mamas.
- Ennis, J. (Writer). (2014). Pay 2 play.
- Ford Foundation. Impact assessment and learning. Retrieved from <http://www.fordfoundation.org/impact/approach>
- Gates Foundation. What we do. Retrieved from <http://www.gatesfoundation.org/what-we-do>
- Green, D., & Patel, M. (2013). *Deepening engagement for lasting impact; a framework for measuring media performance and results*. Retrieved from
- Griffiths, T., Steyvers, M., & Tenenbaum, J. (2007). Topics in semantic representation. *Psychological Review*, 114(2), 211-244.
- Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL*: Morgan Kaufmann.
- Harris, T. A. (Writer). (2014). Through a lens darkly.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America*, 102(46), 16569-16572.
- Jarecki, E. (Writer). (2012). The house I live in.
- John, S., & James, L. (2011). *Impact: a practical guide for evaluating community information projects*. Retrieved from <http://www.knightfoundation.org/publications/impact-practical-guide-evaluating-community-inform>
- Klein, N. (2014). *This changes everything*: Simon & Schuster.
- Knight Foundation. Engaged communities. Retrieved from <http://www.knightfoundation.org/what-we-fund/engaging-communities>
- Knight Foundation. (2011). *Impact: a guide to evaluating community information projects*. Retrieved from
- McCallum, A. K. (2002). MALLET: a machine learning for language toolkit. Retrieved from <http://mallet.cs.umass.edu>
- Media Impact Project. About us. Retrieved from <http://www.mediaimpactproject.org/about.html>
- Napoli, P. (2014). *Measuring media impact: an overview of the field*. Retrieved from
- Piwowar, H. (2013). Altmetrics: value all research products. *Nature*, 493(7431), 159-159.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: a manifesto.
- Reticker, G. (Writer). (2008). Pray the devil back to hell.
- Reticker, G. (Writer). (2011a). I came to testify.
- Reticker, G. (Writer). (2011b). The war we are living.
- Soechtig, S. (Writer). (2014). Fed up.
- Stone, R. (Writer). (2013). Pandora's promise.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS one*, 8(5), e64841.
- Vanclay, F. (1999). Social impact assessment. *Environmental assessment in developing and transitional countries*, 125-135.
- Vanclay, F. (2006). Principles for social impact assessment: a critical comparison between the international and US documents. *Environmental Impact Assessment Review*, 26(1), 3-14.

- Wang, A. (2012). *Women, war & peace. U.S. television broadcast: outreach and audience engagement. impact statement*. Retrieved from
- Whiteman, D. (2004). Out of the theaters and into the streets: a coalition model of the political impact of documentary film and video. *Political Communication*, 21(1), 51-69.
- Woolcock, P. (Writer). (2012). One mile away.